

Temporal Prediction Model of the Evolution of Confirmed Cases of the New Coronavirus (SARS-CoV-2) in Brazil

E. Siqueira, C. Portela, F. Farias, and C. Braga

Abstract—The present article presents a model of temporal prediction about number of confirmed cases evolution of the new coronavirus (SARS-CoV-2) in Brazil. The proposed model is based on the analysis of the data from the source of the Our World in Data platform. The applied data analysis technique was Forecasting (nonlinear regression) implemented in Python programming language. The analysis ranged from 03/14/2020 (100th confirmed case of COVID-19 in Brazil) until 05/15/2020. This model was validated using the cross-validation technique from a comparison of the prediction scenarios and the official numbers released by the Brazilian Health Ministry. The results showed that the proposed model presents a good accuracy for short-term forecasts (7 days).

Index Terms—Data analysis, Forecasting, Mathematical model, Nonlinear equations

I. INTRODUÇÃO

As doenças virais causadas pelos coronavírus tem como alvo principal o sistema respiratório humano. Os surtos anteriores incluem a Síndrome Respiratória Aguda Grave (SARS)-CoV e a Síndrome Respiratória do Oriente Médio (MERS)-CoV [1]. O novo coronavírus, denominado SARS-CoV-2 e causador da doença infecciosa COVID-19, foi detectado inicialmente na cidade de Wuhan, província de Hubei na China, em dezembro de 2019 [2] [3]. A sequência genética do SARS-CoV-2 apresentou mais de 80% de identidade para SARS-CoV e 50% para MERS-CoV [4] [5], sendo que ambos os vírus se originaram a partir de morcegos [6]. Assim, as evidências da análise filogenética indicam que o SARS-CoV-2 pertence ao gênero betacoronavírus, que inclui o SARS-CoV que infecta humanos, morcegos e animais selvagens [7].

A disseminação do SARS-CoV-2 é realizada de pessoa a pessoa e a contaminação ocorre por gotículas respiratórias ou contato, podendo ocorrer de forma continuada. A contração da COVID-19 pode provocar febre, tosse, falta de ar e dificuldades respiratórias [8]. Em alguns casos, as pessoas infectadas podem não apresentar esses sintomas, sendo declaradas assintomáticas. Por outro lado, em casos mais graves, o coronavírus pode ter uma apresentação mais severa, incluindo choque séptico e falência respiratória, podendo levar os infectados à óbito [9].

Elton Siqueira é docente da Faculdade de Sistemas de Informação da Universidade Federal do Pará (UFPA), Cametá, PA, 68400-000 Brasil, email: eltonss@ufpa.br

Carlos Portela e Fabrício Farias são docentes da Faculdade de Sistemas de Informação da Universidade Federal do Pará (UFPA), Cametá, PA, Brasil.

Camila Braga é doutoranda do Programa de Pós-graduação em Biologia Parasitária na Amazônia da Universidade Estadual do Pará, Belém, PA, Brasil.

Seguindo a ordem cronológica dos acontecimentos na China, os primeiros casos da doença ocorreram entre os dias 18 e 29 de dezembro de 2019, com a hospitalização de cinco pacientes apresentando Síndrome do Desconforto Agudo Respiratório (SRDA), sendo que um desses pacientes faleceu [4]. Todos os casos iniciais estavam vinculados ao mercado atacadista de frutos do mar de Huanan, em Wuhan, que comercializa peixes e uma variedade de espécies de animais vivos, incluindo aves, morcegos e cobras [10]. No dia 2 de janeiro de 2020, foram confirmados em laboratório 41 pacientes infectados com a COVID-19 [11]. Em 22 de janeiro, a China já apresentava 571 casos em seu território continental (25 províncias). Três dias depois (25 de janeiro), foram relatados 1975 casos confirmados, contabilizando 56 mortes [5]. No dia 30 de janeiro, foram contabilizados 7734 casos confirmados na China e 90 casos relatados em outros países como Taiwan, Tailândia, Japão, República da Coreia, Emirados Árabes Unidos, Índia, Canadá, França e Alemanha [12]. Relatórios anteriores publicados no final de janeiro de 2020 previram o aparecimento de um potencial surto de coronavírus, dada a estimativa de número básico de reprodução (R_0) ser considerada significativamente maior que 1 (2,24 a 3,58) [13].

No âmbito global, a Organização Mundial da Saúde (OMS) publicou oficialmente a primeira notícia sobre os surtos de doenças do novo coronavírus em 5 de janeiro de 2020 [2]. Em 13 de janeiro, foi confirmado oficialmente na Tailândia o primeiro caso de COVID-19 fora do território chinês. Em 30 de janeiro, após uma viagem de visita técnica para a China, o Comitê de Emergência da OMS resolveu declarar a epidemia uma Emergência de Saúde Pública de Importância Internacional (*Public Health Emergency of International Concern – PHEIC*) e, na tentativa preventiva de conter a proliferação da doença para outras regiões do mundo, publicou um conjunto de orientações técnicas para todos os países [14].

Nas orientações técnicas, a OMS recomendou que todos os países deveriam estar preparados para a contenção da COVID-19 realizando vigilância ativa, detecção precoce, isolamento e gerenciamento de casos, rastreamento de contatos e compartilhamento de dados completos com a OMS [14]. Tal cuidado foi justificado por não existirem vacinas ou tratamentos específicos para a doença, sendo um caminho viável a prevenção contra o contágio e por consequência, contra os efeitos colaterais causados pela infecção.

Embora tenha havido diversos esforços para contenção da COVID-19 por parte da OMS e de governos de diferentes

países, até o dia 24 de maio de 2020 [15], o vírus já estava presente em todos os continentes, com cerca de 5,206.614 milhões de casos confirmados no mundo e 337.736 mil óbitos, sendo o continente americano o mais afetado, com cerca de 2,338.124 milhões de casos. Em relação ao Brasil, o primeiro caso foi confirmado no dia 26 de fevereiro [16] e, em 24 de maio [17], haviam 363.211 mil casos confirmados e 22.666 mil óbitos.

Nesse contexto de pandemia, observa-se um grande esforço da comunidade científica global, em que determinados grupos rapidamente se organizaram para monitorar o número de casos em tempo real, aplicando modelos matemáticos e estatísticos para acompanhar o crescimento do número de infectados por cidade/região/país, permitindo definir estratégias de ação para o combate à proliferação do SARS-CoV-2 [18]. Esses dados são de fundamental importância, pois auxiliam governos e gestores de saúde no planejamento de ações de contenção e atendimento dos doentes [14]. Em relação à disponibilização dos dados e divulgação das informações, destacam-se os painéis (*dashboards*) e relatórios automatizados para o acompanhamento temporal e espacial dos casos confirmados.

No Brasil, destacam-se o Painel Coronavírus do Ministério da Saúde (MS) [17] e o Painel da Rede Covida (PRC) [19], uma iniciativa conjunta entre o Centro de Integração de Dados e Conhecimentos para Saúde (Cidacs/Fiocruz Bahia) e a Universidade Federal da Bahia. Ambas as iniciativas permitem visualizar: os dados atuais divulgados pelas secretarias de saúde dos estados; a evolução do número de casos e óbitos confirmados; a concentração geográfica da doença; e a previsão da evolução dos casos nos próximos dias no Brasil. Em relação aos relatórios automatizados, cita-se o disponibilizado pelo Brasil.io [20], que agrega os dados municipais, os quais não têm sido divulgados pelo MS, considerando a população dos municípios afetados, óbitos suspeitos registrados em cartório, entre outros.

Diante deste contexto, este artigo objetiva contribuir com o combate à COVID-19 apresentando um modelo de predição temporal da evolução do número de casos confirmados do SARS-CoV-2 no Brasil em um horizonte de sete dias. Os dados que alimentam o modelo são coletados em tempo real do sistema *Our World in Data* [21]. Essa análise abrange a partir do centésimo caso confirmado (dia 14/03/2020) até o dia 15/05/2020. Para teste do modelo, estimou-se o número de casos confirmados entre os dias 16 e 22/05/2020. Em seguida, os dados obtidos foram comparados com as predições realizadas pelo PRC [19], sendo ambas confrontadas com os números divulgados pelo MS. Optou-se pela comparação com o PRC devido ao mesmo utilizar aprendizado de máquina supervisionado, assim como o modelo apresentado neste artigo.

Além desta seção introdutória, na Seção II é apresentada a abordagem metodológica para tratar o problema de análise e predições do número de casos da COVID-19 no Brasil. Na Seção III, é descrita a solução proposta através da definição e treinamento do modelo proposto. Em seguida, a Seção IV discorre sobre os testes de validação e a aplicação do modelo de predição. Por fim, a Seção V apresenta as principais conclusões desta pesquisa.

II. ABORDAGEM DO PROBLEMA

A fim de propor um modelo de predição eficiente, foram realizadas as seguintes etapas: i) identificação e verificação de um repositório de dados que fornecesse dados confiáveis e estruturados para definir um modelo consistente no que diz respeito a realidade do Brasil. Nesse sentido, utilizou-se a base de dados do *Our World in Data* [21]. Trata-se de uma plataforma digital especializada em expor pesquisas empíricas sobre mudanças nas condições de qualidade de vida e suas consequências no mundo através de visualizações de dados interativas, principalmente gráficos e mapas; ii) tratamento dos dados e definição de alguns parâmetros com intuito de ajustar o modelo em questão; iii) implementação do modelo de predição temporal (*Forecasting*) [22] (regressão não-linear) de modo a estimar o número de casos confirmados no Brasil; iv) validação do desempenho desse modelo através da técnica de *cross-validation* com previsões simuladas [23]; e v) realização de predições, a partir do dia 16/05/2020, visando apresentar cenários para os próximos 7 dias.

Na etapa i, constatou-se que os dados do *Our World in Data* selecionados para definição do modelo de predição estão estruturados em: *iso_code*, *location*, *date*, *total_cases*, *new_cases*, *new_deaths*, *total_death*, *total_cases_per_million*, *total_deaths_per_million*, *new_deaths_per_million*, *total_tests*, *new_tests*, *total_tests_per_thousand*, *new_tests_per_thousand* e *tests_units*. Para definição do modelo, foram considerados apenas os atributos *date* (data), *location* (localização) e casos confirmados (*new_cases*). A partir da análise desses dados, observou-se o padrão de crescimento do número de casos confirmados. No dia 26/02/2020, houve o primeiro caso de infecção registrado pelo MS e, após 3 semanas (18/03/2020) houve um salto para 428 casos confirmados, sendo resultado de um aumento significativo da velocidade de disseminação da COVID-19 no Brasil. Em 21/03/2020, foi registrada uma quantidade acima de 1000 casos confirmados e no dia 28/03/2020 (uma semana após o milésimo caso), essa marca quase quadruplicou, chegando aproximadamente a 3900 casos confirmados. Na Figura 1 é apresentada a evolução dos casos confirmados no Brasil no período de 26/02/2020 até 15/05/2020.

Com base na Figura 1, é possível verificar que o crescimento da dispersão da COVID-19 no Brasil está seguindo o mesmo padrão de comportamento dos países mais afetados pelo vírus, como por exemplo, os Estados Unidos, cuja evolução do número de casos, entre 20/01/2020 (primeiro caso confirmado) [24] e 15/05/2020, é apresentada na Figura 2.

Uma pesquisa com objetivo semelhante a apresentada neste artigo foi realizada por Arenas et al. (2020) [25], na qual foi utilizada uma adaptação do modelo de mobilidade baseado na cadeia de Markov para dispersões epidêmicas em redes complexas. No entanto, essa pesquisa foi aplicada na Espanha, usando vários parâmetros epidemiológicos fornecidos pelo *National Institute of Statistics*, como informações sobre as unidades de tratamento intensivo, dados da mobilidade urbana, taxa de morte, dentre outros parâmetros. Adicionalmente, esse modelo levou em consideração a dinâmica de dispersão da COVID-19, parametrizando: i) o efeito das pessoas assintomáticas, pois elas podem expandir o nível de contaminação;

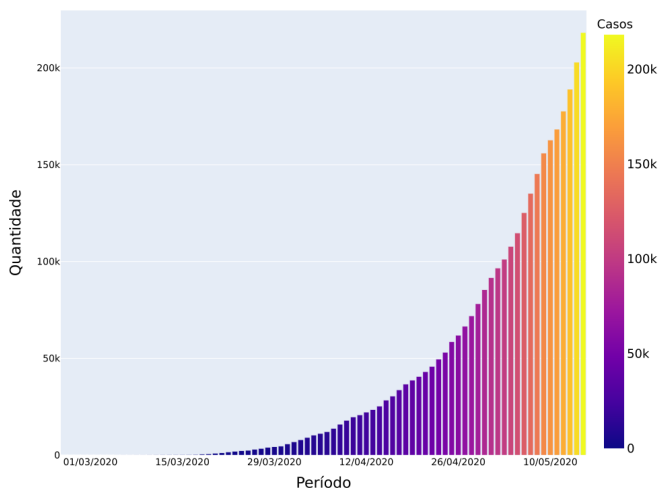


Fig. 1. Evolução do número de casos confirmados da COVID-19 no Brasil. Fonte: [21]

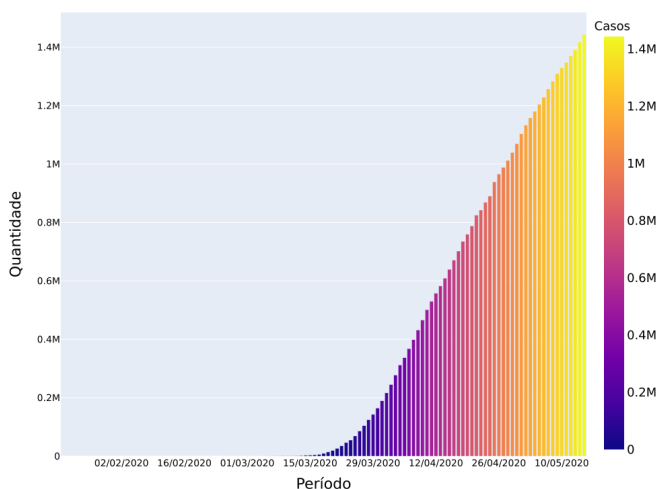


Fig. 2. Evolução do número de casos confirmados da COVID-19 nos Estados Unidos. Fonte: [21]

ii) a fração dos indivíduos que exigem hospitalização nas unidades de tratamento intensivo. Os resultados indicaram que, na ausência de restrições de mobilidade, o pico de incidências da contaminação na Espanha ocorreria na metade do mês de abril de 2020.

Os trabalhos de Yang *et al.* [26] e Anastassopoulou *et al.* [27] destacam o massivo compartilhamento de dados na China em prol do combate da doença. No entanto, no Brasil os dados disponibilizados pelo Governo Federal são restritos aos casos graves e óbitos, isto é, ainda há carência de maior divulgação sobre os parâmetros epidemiológicos, cumprimento do distanciamento social, número de testes realizados, efeitos da propagação do vírus, entre outros. Deste modo, inviabilizando a construção de um modelo de predição mais complexo.

Embora exista a limitação de dados, o trabalho proposto pretende contornar essa problemática criando um modelo que

utiliza dois dados relevantes como entrada (data e número de casos confirmados), de modo que os parâmetros do modelo sejam ajustados durante a etapa de treinamento. Além disso, a abordagem proposta pode ser reproduzida em uma escala menor, isto é, municipal, estadual e regional, assim possibilitando ações estratégicas para combate à COVID-19 pelos gestores.

III. SOLUÇÃO PROPOSTA

O modelo de predição proposto neste trabalho objetiva analisar os dados da contaminação do novo coronavírus representado na Figura 1. Assim, na etapa ii, conforme constatado o crescimento exponencial da contaminação do vírus, adotou-se uma função logística e, a partir das simulações feitas, encontrou-se um conjunto de parâmetros adequados ao modelo que correspondem ao ponto de saturação (sinalização dos efeitos das políticas de contenção sobre a circulação de pessoas e o ponto de achatamento da curva) e os intervalos de incerteza que têm por finalidade ajustar o modelo para futuras mudanças de tendência dos dados, considerando que o futuro sofrerá mudanças de tendência semelhantes às da história. Em particular, assume-se que a frequência média das mudanças de tendência no futuro serão as mesmas que se observou nos outros países. Logo, projetam-se as mudanças de tendência por meio dos intervalos de incerteza [22].

Na etapa iii, utilizou-se a linguagem de programação Python, com as bibliotecas Prophet (<https://facebook.github.io/prophet/>) e Plotly (<https://pypi.org/project/plotly/>), a fim de implementar algorítmicamente o modelo e gerar os gráficos de predições. Nesta implementação, foi adotado o modelo para predição de séries temporais [28]. Este modelo utiliza três componentes principais: tendência, sazonalidade e feriados, combinadas conforme a Equação 1:

$$y(t) = g(t) + s(t) + h(t) + \epsilon \quad (1)$$

O componente $g(t)$ é a função de tendência que modela mudanças não periódicas no valor da série temporal, $s(t)$ representa mudanças periódicas (por exemplo, sazonalidade semanal e anual), e $h(t)$ representa os efeitos de feriados que ocorrem em horários potencialmente irregulares de um ou mais dias. O termo (ϵ) representa quaisquer alterações idiossincráticas (situações incomuns) que não são acomodadas pelo modelo. Além disso, para o cenário epidêmico em questão, considerou-se o modelo no qual há crescimento não linear que satura em uma capacidade de carga, e geralmente, esse comportamento é modelado usando o modelo de crescimento logístico dado pela Equação 2:

$$g(t) = \frac{C}{1 + \exp(-k(t - m))} \quad (2)$$

Nesta fórmula, C é a capacidade de carga, k a taxa de crescimento e m um parâmetro de deslocamento (*offset*). O estudo de Buckingham-Jeffery *et al.* [29] mostra o impacto negativo dos feriados nas previsões do modelo, uma vez que esses pontos são tidos como independentes e imprevisíveis, levando a resultados errôneos pelo sistema de vigilância de saúde. Visto isso, o presente modelo busca contornar esse viés causado pelos feriados removendo o termo $h(t)$ da

Equação 1, tratando os efeitos dos feriados como eventos não independentes e iguais em todas as janelas de tempo. Assim, os efeitos destes eventos não impactam na previsão do cenário epidêmico em questão.

Na etapa iv, a fim de validar o desempenho do modelo proposto, utilizou-se a técnica de *cross-validation* [22] com simulações adotando um horizonte de 14 dias. Mais especificamente, utilizou-se a métrica chamada Erro Médio Absoluto Percentual (*Mean Absolute Percentage Error – MAPE*), pois é uma medida de precisão frequentemente usada em modelos de regressão e em problemas relacionados a estimativa de tendências [23] [30] [31]. Por fim, na etapa v, realizaram-se as previsões considerando um limite de saturação de 2 milhões de casos confirmados e um período de estimativa de 7 dias. As etapas iv e v da metodologia abrangem os principais achados desta pesquisa e são descritas em mais detalhes na próxima seção.

IV. RESULTADOS

Após a definição do modelo preliminar de previsão, realizou-se o treinamento do mesmo que depende diretamente da quantidade e qualidade dos dados disponíveis. Neste sentido, quanto mais dados disponíveis em um maior intervalo de tempo, melhor será o processo de treinamento do modelo e, conseqüentemente, melhor serão as estimativas geradas por ele. É válido reforçar que para este trabalho foram considerados para alimentação do modelo os dados provenientes da plataforma *Our World in Data*, a partir da data de notificação do paciente 100 (dia 14/03/2020) até 15 de Maio de 2020 ($N = 62$ dias). Nesta etapa de treinamento do modelo, os parâmetros C , k e m da Equação 2 foram definidos da seguinte maneira: i) para capacidade de carga, usamos $C = 4000000$ (ponto de saturação do número de casos confirmados) que foi extraído através das diversas execuções do experimento conforme a diminuição da taxa de erro; ii) O conjunto de valores de k e m são estimados por amostragem bayesiana [32] [33]. Utilizando esses dados de entrada, o modelo foi treinado e obteve o ajuste apresentado na Figura 3, respeitando os intervalos de incerteza de 95%.

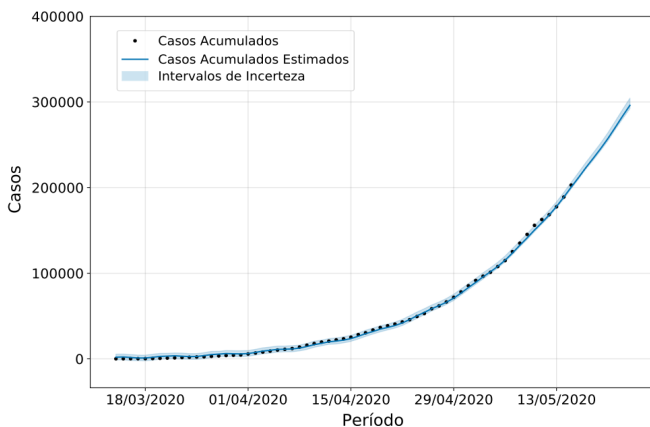


Fig. 3. Treinamento do modelo de previsão proposto.

Na etapa de validação, observou-se que as margens de erro do modelo ficam em torno de 5.9% para previsões com

horizonte de 14 dias e 1% para horizontes de 7 dias, ou seja, o modelo apresentou uma boa acurácia durante as simulações realizadas, uma vez que os erros são relativamente baixos, conforme mostra o gráfico da Figura 4.

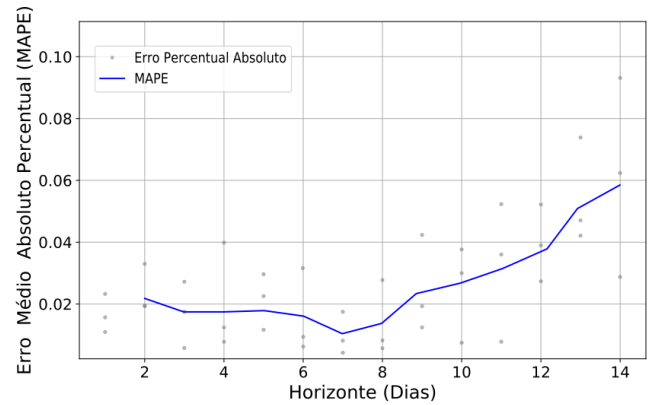


Fig. 4. Validação do modelo proposto usando Erro Médio Absoluto Percentual.

O estudo conduzido por Fanelli e Piazza [34] salienta que as ações como o *lockdown* podem resultar em respostas estratégicas aplicadas em um curto espaço de tempo, como de uma semana. Deste modo, desencorajando a utilização de previsões de longos períodos para prever o comportamento da pandemia, visto que o modelo proposto busca apoiar as autoridades na tomada de decisão imediata para o combate à COVID-19 considerando o crescimento exponencial do número de casos no Brasil. Diante disso e com base nas informações obtidas na etapa de validação, o presente trabalho optou por realizar as previsões com horizontes de 7 dias por dois motivos: i) Desempenho: os melhores resultados no que diz respeito a acurácia são alcançados para horizontes dessa periodicidade (veja a Figura 4); ii) Saúde Pública: permite diminuir significativamente o impacto da doença, pois os governos podem alterar suas estratégias de ação de acordo com a dispersão do vírus e podem planejar com antecedência as etapas preventivas necessárias, como campanhas de saúde pública e conscientização dos cidadãos. Ressalta-se que o modelo desenvolvido pode realizar previsões de horizontes maiores com propósitos de adoção de políticas públicas a longo prazo, entretanto a taxa de erro tende a aumentar para esses horizontes maiores.

Após a etapa de treinamento e validação do modelo, foi realizada a comparação entre o modelo proposto com os dados fornecidos pelo MS. Além disso, a comparação foi estendida com os dados fornecidos pelo modelo de previsão do PRC [19], conforme apresenta a Tabela I. A partir desses dados, selecionaram-se duas medidas comumente utilizadas na literatura [35] [31] no que diz respeito a comparação de desempenho de modelos de previsão temporal: i) Raiz do Erro Quadrático Médio (*Root Mean Squared Error – RMSE*), a fim de medir o erro médio do modelo na previsão do resultado de uma observação; ii) Erro Médio Absoluto Percentual (*Mean Absolute Percentage Error – MAPE*) com o objetivo de medir o tamanho do erro absoluto em termos

de porcentagem. Esses indicadores foram utilizados de modo a comparar o desempenho entre os modelos com os dados divulgados pelo MS. Além disso, com o intuito de facilitar a interpretação e comparação entre os modelos dentro de uma escala de $[0, 1]$, utilizou-se a forma normalizada do RMSE, chamada de *Normalized Root Mean Squared Error* (NRMSE).

TABELA I

COMPARAÇÃO ENTRE OS DADOS OBSERVADOS PELO MINISTÉRIO DA SAÚDE E OS VALORES ESTIMADOS (PREDIÇÕES) PELO PRC E MODELO PROPOSTO

Data	Dados	Dados	Erro	Dados	Erro
	Observados (MS)	Estimados (PRC)	Residual (MS - PRC)	Estimados (Modelo)	Residual (MS - Modelo)
16/05/2020	233142	230586	2556	228657	4485
17/05/2020	241080	245099	-4019	242417	-1337
18/05/2020	254220	260495	-6275	257000	-2780
19/05/2020	271628	276827	-5199	273125	-1497
20/05/2020	291579	294151	-2572	290817	762
21/05/2020	310087	312527	-2440	309725	362
22/05/2020	330890	332020	-1130	329596	1294
MAPE			1.31		0.7
RMSE			3834.74		2212.25
NRMSE			0.039		0.023

A partir da análise dos dados apresentados na Tabela I, observa-se que os valores do MAPE, RMSE e NRMSE do modelo proposto foram menores que os valores indicados pelo modelo do PRC. Conforme as técnicas de análise de dados sobre modelos de predição temporal [35], quanto menor esses indicadores, melhor a predição do modelo. Nesse sentido, o modelo proposto neste artigo apresentou um melhor desempenho em comparação às predições do modelo do PRC [19].

No entanto, a subnotificação dos casos é uma limitação dos dados apresentados pelo MS, pois os testes da COVID-19 no Brasil são realizados prioritariamente em profissionais de saúde ou de segurança, casos mais graves (pacientes hospitalizados) e óbitos [36]. Ainda assim, nenhum desses casos é garantia de exame, devido à escassez de kits de testes no país. Segundo Roser *et al.* (2020) [21], a contagem de casos confirmados depende de quanto um país consegue testar sua população para entender a dinâmica da disseminação da doença no seu território. Sem esses dados não é possível saber quais estados estão tomando as medidas de contenção adequadas e quais estão subnotificando casos e óbitos decorrentes da COVID-19. Neste contexto, o MS não divulga quantos testes para coronavírus já foram realizados no Brasil [36]. No entanto, estima-se que a proporção de testes por cada 1 mil habitantes seja de apenas 0.63 (ou 63 por cada 100 mil habitantes), considerando uma população de 210 milhões de habitantes [37]. Essa taxa é inferior à de outros países latino-americanos, como Chile (6.43) e Cuba (2.65).

Destaca-se que esse modelo não considera o atraso nas notificações e nem a subnotificação dos casos existentes nos dados disponibilizados pelo MS. Entretanto, recomenda-se a abordagem de Bastos *et al.* (2019) [38] que propõe uma modelagem hierárquica bayesiana como forma de corrigir os atrasos dos relatórios de notificações e quantificar a incerteza associada às predições.

Apesar dessas limitações, um diferencial do modelo pro-

posto é a utilização de dados em tempo real, sendo os parâmetros ajustados durante a etapa de treinamento. Enquanto que os demais modelos de predições propostos na literatura [19], [26] e [27] utilizam parâmetros estáticos que são derivados de observações e estudos epidemiológicos. Embora os estudos com parâmetros estáticos tenham a devida validade científica, estes sofrem influências a partir do ambiente, região e local da análise, assim podendo inviabilizar a reprodução desses modelos em outros cenários. Além disso, o modelo proposto pode ser reproduzível em outras escalas geográficas (municipal, estadual e regional) de modo a não necessitar do conhecimento prévio dos parâmetros epidemiológicos que os modelos anteriormente citados requerem para realizar as predições. Por exemplo, o modelo do PRC [19] utiliza parâmetros epidemiológicos derivados do modelo SEIR, que são dependentes da população, taxa de infecção e outros fatores externos, logo esse modelo não pode acompanhar os diferentes comportamentos da dispersão do vírus nos diferentes cenários apresentados no Brasil. Desta forma, ao se utilizar as predições do modelo do PRC seria necessário um novo estudo e observações complementares para uma nova estimativa desses parâmetros a fim de se adequar ao comportamento da curva epidemiológica.

V. CONCLUSÕES

Nesta fase avançada de transmissão da COVID-19, a quantidade de Unidades de Terapia Intensiva (UTIs) disponíveis são insuficientes para o atendimento pleno da população. Essa saturação já é observada em alguns estados brasileiros, como no Amazonas, Ceará, Pará e Pernambuco que têm ocupação de mais de 90% dos leitos de UTI para COVID-19 [39]. Para exemplificar esse cenário, o Cadastro Nacional de Estabelecimentos de Saúde do SUS informa que a cidade de Manaus, capital do Amazonas, possui 273 leitos de UTI [40] e o governo do estado divulgou no dia 22 de abril de 2020 que 269 leitos já estavam ocupados [41], representando um total de 98.5% de ocupação. No caso do Brasil, vale ressaltar que esta investigação visa alertar as autoridades para que sejam realizados os esforços necessários para evitar o aumento do número de casos, conforme previsto pelo modelo proposto, pois a maioria das cidades brasileiras possuem poucos leitos de UTI e baixo número de equipamentos médicos. Deste modo, combater o acelerado avanço do número de casos é tarefa crucial nesta crise, assim evitando a proliferação de casos graves que podem evoluir para uma síndrome respiratória aguda grave, precisando de leitos de UTI que o país não dispõe.

Por essa razão, nos momentos em que a saúde pública enfrenta desafios, como uma pandemia, o governo precisa articular planos de contingência para evitar a propagação da doença, criando políticas públicas efetivas, com base em previsões sobre as taxas de contaminação na população [14]. Administradores de hospitais, governos e formuladores de políticas públicas devem se preparar para um aumento substancial na capacidade de leitos de cuidados intensivos [42]. Neste sentido, os dados provenientes de trabalhos de predição são essenciais para projetos de destinação de bens e recursos nos diferentes estados brasileiros.

Em modelos epidemiológicos, quanto maior o número de dados, melhor será sua acurácia [43]. Portanto, é importante realizar as previsões com horizontes de curto e médio prazo, como os cenários de 7 dias previstos no modelo apresentado neste artigo. Desta forma, os gestores públicos podem ter uma visão rápida e com baixo índice de erros da magnitude do problema, para assim, de forma correta e embasada, implementar as políticas públicas buscando minimizar o impacto da disseminação da COVID-19.

Os resultados apresentados neste trabalho serão ajustados com as próximas atualizações da plataforma *Our World in Data*. Os autores continuarão com os testes de validação, definindo novos parâmetros para o modelo a fim de atingir uma maior acurácia para as previsões. Para a evolução do modelo, pretende-se, como trabalho futuro, implementar a abordagem de Bastos *et al.* (2019) [38] para correção dos atrasos nas notificações e subnotificações de casos.

Além disso, é importante destacar que muitos parâmetros não disponibilizados pelo Governo Federal podem impactar nas previsões do modelo, como a taxa da população brasileira que está em quarentena, a efetividade dessa quarentena, a abrangência do número de testes realizados na população, o monitoramento dos pacientes com contaminação confirmada e das pessoas que tiveram contato com esses, dentre outros.

Por fim, pode-se concluir que o modelo de previsão proposto se ajustou ao comportamento da curva de contágio do Brasil, tornando possível reproduzir o modelo em um escala menor (regional, estadual ou municipal) usando como entrada de dados a data de ocorrência e número de casos confirmados. Em termos práticos, a adoção de um modelo que prediz os resultados a partir dos dados reais de contaminação é uma alternativa que possibilita uma maior consistência no processo de tomada de decisões estratégicas para o combate à COVID-19, assim como, pode servir para a redução do impacto da pandemia, principalmente nos setores da saúde e economia. No setor da saúde, ao se observar um aumento no número de casos numa determinada semana, pode-se intensificar campanhas de conscientização, realizar testes rápidos para triagem e encaminhamento de pacientes, por exemplo encaminhar casos mais graves para hospitais de campanha, a fim de não sobrecarregar os leitos dos hospitais públicos que além da COVID-19, tratam de outras doenças. Já no setor da economia, em caso do aumento do número de casos em um curto período de tempo, pode-se determinar o início de um período de *lockdown* ou decidir pelo fechamento de setores não essenciais, como por exemplo bares, academias e salões de beleza.

AGRADECIMENTOS

Ao Ministério da Saúde pela disponibilização e atualização do Painel Coronavírus. Ao Centro de Integração de Dados e Conhecimentos para Saúde (Cidacs/Fiocruz Bahia) e a Universidade Federal da Bahia pela disponibilização do Painel da Rede Covida. Aos responsáveis pela disponibilização e análise de dados do *Our World in Data*.

REFERÊNCIAS

- [1] I. I. Bogoch, A. Watts, A. Thomas-Bachli, C. Huber, M. U. G. Kraemer, and K. Khan, "Pneumonia of unknown aetiology in wuhan, china: potential for international spread via commercial air travel," *Journal of travel medicine*, vol. 27, no. 2, March 2020. [Online]. Available: <https://europepmc.org/articles/PMC7107534>
- [2] W. H. Organization, "Who timeline - covid-19," <https://www.who.int/news-room/detail/08-04-2020-who-timeline---covid-19>, April 2020, (Accessed on 05/24/2020).
- [3] A. Du Toit, "Outbreak of a novel coronavirus," *Nature Reviews Microbiology*, vol. 18, no. 3, p. 123, 2020. [Online]. Available: <https://doi.org/10.1038/s41579-020-0332-0>
- [4] L.-L. Ren, Y. Wang, Z.-Q. Wu, Z.-C. Xiang, L. Guo, T. Xu, Y.-Z. Jiang, Y. Xiong, Y.-J. Li, H. Li, G.-H. Fan, X.-Y. Gu, Y. Xiao, H. Gao, J.-Y. Xu, F. Yang, X.-M. Wang, C. Wu, L. Chen, and J. Wang, "Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study," *Chinese medical journal*, vol. 133, 01 2020.
- [5] W. Wang, J. Tang, and F. Wei, "Updated understanding of the outbreak of 2019 novel coronavirus (2019-ncov) in wuhan, china," *Journal of Medical Virology*, vol. 92, no. 4, pp. 441–447, 2020. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.25689>
- [6] J. Cui, F. Li, and Z.-L. Shi, "Origin and evolution of pathogenic coronaviruses," *Nature Reviews Microbiology*, vol. 17, no. 3, pp. 181–192, 2019. [Online]. Available: <https://doi.org/10.1038/s41579-018-0118-9>
- [7] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G. F. Gao, and W. Tan, "A novel coronavirus from patients with pneumonia in china, 2019," *New England Journal of Medicine*, vol. 382, no. 8, pp. 727–733, 2020, pMID: 31978945. [Online]. Available: <https://doi.org/10.1056/NEJMoa2001017>
- [8] M. da Saúde, "O que é coronavírus (covid-19)," <https://coronavirus.saude.gov.br/>, April 2020, (Accessed on 05/26/2020).
- [9] "A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster," *The Lancet*, vol. 395, no. 10223, Feb. 2020.
- [10] W. H. Organization, "Who director-general's remarks at the media briefing on 2019-ncov on 11 february 2020," <https://www.who.int/dg/speeches/detail/who-director-general-s-remarks-at-the-media-briefing-on-2019-ncov-on-11-february-2020>, February 2020, (Accessed on 05/26/2020).
- [11] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, and B. Cao, "Clinical features of patients infected with 2019 novel coronavirus in wuhan, china," *The Lancet*, vol. 395, 01 2020.
- [12] M. Bassetti, A. Vena, and D. R. Giacobbe, "The novel chinese coronavirus (2019-ncov) infections: Challenges for fighting the storm," *European Journal of Clinical Investigation*, vol. 50, no. 3, p. e13209, 2020. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/eci.13209>
- [13] S. Zhao, Q. Lin, J. Ran, S. S. Musa, G. Yang, W. Wang, Y. Lou, D. Gao, L. Yang, D. He, and M. H. Wang, "Preliminary estimation of the basic reproduction number of novel coronavirus (2019-ncov) in china, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak," *International Journal of Infectious Diseases*, vol. 92, pp. 214 – 217, 2020. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1201971220300539>
- [14] W. H. Organization, "Health systems respond to covid-19 - technical guidance," http://www.euro.who.int/__data/assets/pdf_file/0006/437469/TG2-CreatingSurgeAcuteUCCapacity-eng.pdf?ua=1, April 2020, (Accessed on 05/24/2020).
- [15] —, "Coronavirus disease (covid-19) outbreak situation," <https://covid19.who.int/>, May 2020, (Accessed on 05/24/2020).
- [16] M. da Saúde, "Brasil confirma primeiro caso da doença," <https://www.saude.gov.br/noticias/agencia-saude/46435-brasil-confirma-primeiro-caso-de-novo-coronavirus>, February 2020, (Accessed on 05/24/2020).
- [17] —, "Painel coronavírus," <https://covid.saude.gov.br/>, 2020 February, (Accessed on 05/24/2020).
- [18] R. M. Lana, F. C. Coelho, M. F. d. C. Gomes, O. G. Cruz, L. S. Bastos, D. A. M. Villela, and C. T. Codeço, "Emergência do novo coronavírus (SARS-CoV-2) e o papel de uma vigilância nacional em saúde oportuna e efetiva," vol. 36, 2020.
- [19] R. CoVida, "Painel corona vírus," <https://painel.covid19br.org/>, March 2020, (Accessed on 05/24/2020).
- [20] Brasil.IO, "Especial covid-19 - dados por município," <https://brasil.io/covid19/>, March 2020, (Accessed on 05/24/2020).
- [21] E. O.-O. Max Roser, Hannah Ritchie and J. Hasell, "Coronavirus pandemic (covid-19)," *Our World in Data*, 2020, <https://ourworldindata.org/coronavirus>.

[1] I. I. Bogoch, A. Watts, A. Thomas-Bachli, C. Huber, M. U. G. Kraemer, and K. Khan, "Pneumonia of unknown aetiology in wuhan,

- [22] S. J. Taylor and B. Letham, "Forecasting at scale," *PeerJ PrePrints*, vol. 5, p. e3190, 2017.
- [23] J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*, ser. The Morgan Kaufmann Series in Data Management Systems. Elsevier Science, 2011. [Online]. Available: <https://books.google.com.br/books?id=pQws07dpjoc>
- [24] M. L. Holshue, C. DeBolt, S. Lindquist, K. H. Lofy, J. Wiesman, H. Bruce, C. Spitters, K. Ericson, S. Wilkerson, A. Tural, G. Diaz, A. Cohn, L. Fox, A. Patel, S. I. Gerber, L. Kim, S. Tong, X. Lu, S. Lindstrom, M. A. Pallansch, W. C. Weldon, H. M. Biggs, T. M. Uyeki, and S. K. Pillai, "First case of 2019 novel coronavirus in the united states," *New England Journal of Medicine*, vol. 382, no. 10, pp. 929–936, 2020, pMID: 32004427. [Online]. Available: <https://doi.org/10.1056/NEJMoa2001191>
- [25] A. Arenas, W. Cota, J. Gomez-Gardenes, S. Gómez, C. Granell, J. T. Matamalas, D. Soriano-Panos, and B. Steinegger, "A mathematical model for the spatiotemporal epidemic spreading of covid19," *medRxiv*, 2020. [Online]. Available: <https://www.medrxiv.org/content/early/2020/03/23/2020.03.21.20040022>
- [26] Z. Yang, Z. Zeng, K. Wang, S.-S. Wong, W. Liang, M. Zanin, P. Liu, X. Cao, Z. Gao, Z. Mai, J. Liang, X. Liu, S. Li, Y. Li, F. Ye, W. Guan, Y. Yang, F. Li, S. Luo, Y. Xie, B. Liu, Z. Wang, S. Zhang, Y. Wang, N. Zhong, and J. He, "Modified seir and ai prediction of the epidemics trend of covid-19 in china under public health interventions," *Journal of Thoracic Disease*, vol. 12, no. 3, 2020. [Online]. Available: <http://jtd.amegroups.com/article/view/36385>
- [27] C. Anastassopoulou, L. Russo, A. Tsakris, and C. Siettos, "Data-based analysis, modelling and forecasting of the covid-19 outbreak," *PLOS ONE*, vol. 15, no. 3, pp. 1–21, 03 2020. [Online]. Available: <https://doi.org/10.1371/journal.pone.0230405>
- [28] A. C. Harvey and S. Peters, "Estimation procedures for structural time series models," *Journal of Forecasting*, vol. 9, no. 2, pp. 89–108, 1990. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/for.3980090203>
- [29] E. Buckingham-Jeffery, R. Morbey, T. House, A. J. Elliot, S. Harcourt, and G. E. Smith, "Correcting for day of the week and public holiday effects: improving a national daily syndromic surveillance service for detecting public health threats," *BMC Public Health*, vol. 17, no. 1, p. 477, 2017. [Online]. Available: <https://doi.org/10.1186/s12889-017-4372-y>
- [30] S. Wang and W. Chaovalitwongse, *Evaluating and Comparing Forecasting Models*, 02 2011.
- [31] j. Jiménez, K. Donado, and C. G. Quintero, "A methodology for short-term load forecasting," *IEEE Latin America Transactions*, vol. 15, no. 3, pp. 400–407, 2017.
- [32] Y. Zhao, J. Zhang, and C. Chen, "Self-adversarially learned bayesian sampling," *ArXiv*, vol. abs/1811.08929, 2019.
- [33] C. Rodriguez Rivero, J. Pucheta, M. Herrera, V. Sauchelli, and S. Laboret, "Time series forecasting using bayesian method: Application to cumulative rainfall," *IEEE Latin America Transactions*, vol. 11, no. 1, pp. 359–364, 2013.
- [34] D. Fanelli and F. Piazza, "Analysis and forecast of COVID-19 spreading in China, Italy and France," *Chaos, Solitons and Fractals*, vol. 134, p. 109761, 2020. [Online]. Available: <https://doi.org/10.1016/j.chaos.2020.109761>
- [35] A. Botchkarev, "Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology," pp. 1–37, 2018. [Online]. Available: <http://arxiv.org/abs/1809.03006>
- [36] P. Watanabe, "Mesmo falta de ar não garante testes para novo coronavírus no brasil - 01/04/2020 - equilíbrio e saúde - folha," <https://www1.folha.uol.com.br/equilibrioesaude/2020/04/mesmo-falta-de-ar-nao-garante-testes-para-novo-coronavirus-no-brasil.shtml>, April 2020, (Accessed on 05/24/2020).
- [37] L. Barrucho, "Brasil é um dos países que menos realiza testes para covid-19, abaixo de cuba e chile - bbc news brasil," <https://www.bbc.com/portuguese/internacional-52383539>, April 2020, (Accessed on 05/20/2020).
- [38] L. S. Bastos, T. Economou, M. F. C. Gomes, D. A. M. Villela, F. C. Coelho, O. G. Cruz, O. Stoner, T. Bailey, and C. T. Codeço, "A modelling approach for correcting reporting delays in disease surveillance data," *Statistics in Medicine*, vol. 38, no. 22, pp. 4363–4377, 2019. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.8303>
- [39] B. Lima, "Quatro estados têm ocupação máxima de utis e preocupam governo - brasil," <https://www.correiobraziliense.com.br/app/noticia/brazil/2020/04/20/interna-brasil,846333/quatro-estados-tem-ocupacao-maxima-de-utis-e-preocupam-governo.shtml>, April 2020, (Accessed on 05/20/2020).
- [40] DataSUS, "Cnesweb - cadastro nacional de estabelecimentos de saúde," http://cnes2.datasus.gov.br/Mod_Ind_Tipo_Leito.asp?VEstado=13&VMun=130400, 2020, (Accessed on 05/20/2020).
- [41] F. de Vigilância em Saúde do Amazonas, "Amazonas chega a 2479 casos de covid-19 e governo amplia parcerias para mais leitos," http://www.fvs.am.gov.br/noticias_view/3865, April 2020, (Accessed on 05/20/2020).
- [42] J. Phua, L. Weng, L. Ling, M. Egi, C.-M. Lim, J. V. Divatia, B. R. Shrestha, Y. M. Arabi, J. Ng, C. D. Gomersall, M. Nishimura, Y. Koh, and B. Du, "Intensive care management of coronavirus disease 2019 (COVID-19): challenges and recommendations," *The Lancet Respiratory Medicine*, vol. 8, no. 5, pp. 506–517, may 2020. [Online]. Available: [https://doi.org/10.1016/S2213-2600\(20\)30161-2](https://doi.org/10.1016/S2213-2600(20)30161-2)
- [43] R. Cabral, S. Chase, R. Ribeiro, G. Marques, E. Morais, A. Zissou, E. Andrade, W. Couto, P. Júnior, P. Campos, J. Almeida, and O. Chase, "Estudo inicial sobre a evolução do novo coronavírus (sars-cov-2) no estado do Pará (brasil), no período entre 17/03/2020 e 06/04/2020," *Brazilian Journal of Health Review*, vol. 3, pp. 2914–2931, 01 2020.



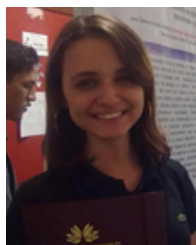
Elton Siqueira Doutorando em Informática na Universidade de Brasília (UnB), possui bacharelado em Sistemas de Informação e mestrado em Ciência da Computação pela Universidade Federal do Pará (UFPA) e Licenciatura em Física pelo Instituto Federal do Pará. Atualmente é Professor Adjunto da UFPA, atuando na Faculdade de Sistemas de Informação, Campus Cametá.



Carlos Portela Doutor em Ciência da Computação pela Universidade Federal de Pernambuco (UFPE), possui mestrado em Ciência da Computação pelo Centro de Informática (CIn) da UFPE e bacharelado em Sistemas de Informação pela Universidade Federal do Pará (UFPA). Atualmente é Professor Adjunto da UFPA, atuando na Faculdade de Sistemas de Informação, Campus Cametá.



Fabricio Farias Doutor e Mestre em Engenharia Elétrica pela Universidade Federal do Pará (UFPA), possui graduação em Engenharia da Computação pela UFPA. Atualmente é Professor Adjunto da UFPA, atuando na Faculdade de Sistemas de Informação, Campus Cametá.



Camila Braga Doutoranda em Biologia Parasitária na Amazônia da Universidade do Estado do Pará (UEPA), possui mestrado em Biologia Parasitária na Amazônia pela UEPA, tendo experiência na área de Microbiologia, atuando principalmente nos seguintes temas: Arbovírus, Bioinformática e Metagenômica.